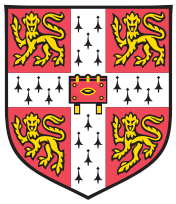
Computer Laboratory
University of Cambridge



Computer Laboratory
University of Cambridge
Cambridge, UK, CB3 0FD

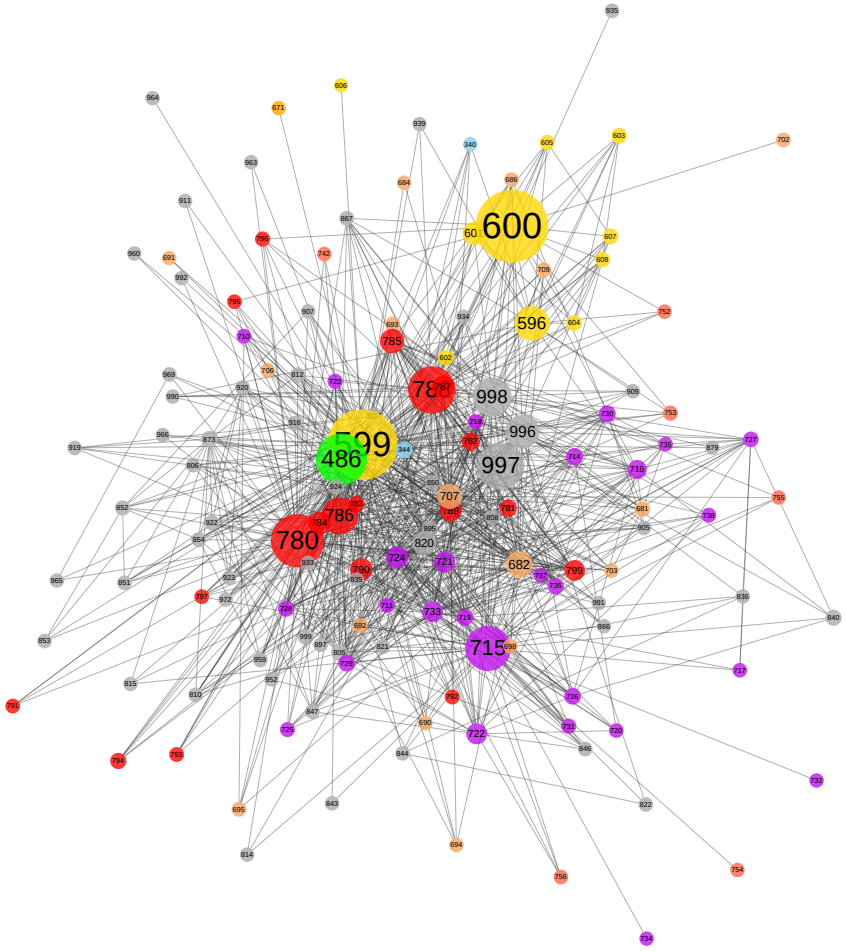
User Manual

CytoCom: a Cytoscape plugin to visualise, query and analyse disease-disease dynamic networks

CytoCom, Version 1.0.0

Haoming Xu, Mohammad Ali Moni and Pietro Liò

October 2014



User Manual

October 2014

CytoCom: a Cytoscape plugin to visualise, query and analyse disease-disease dynamic networks

CytoCom, Version 1.0.0

Haoming Xu, Mohammad Ali Moni and Pietro Liò

Computer Laboratory
University of Cambridge
Cambridge
UK, CB3 0FD

Table of Contents

System Requirements

In order to use CytoCom, your computer must be equipped with the following software package:

- Java 1.6+ is needed, Java 1.7 is recommended.
- Cytoscape 3.1.0 + installed

To download CytoCom, please visit: `http://www.cl.cam.ac.uk/~mam211/`. In Cytoscape 3.1.0 or later version, the App Manager allows users to quickly install and uninstall the apps. After downloading CytoCom, please go to Apps→ app Manager and then click the "Install from File" button at the bottom of the "Install Apps" tab. Select the downloaded CytoCom.jar to install. After installing CytoCom, it could be found under the "Control Panel".

Documentation

Understanding the comorbidity of human diseases is one of the most challenging issues in bioinformatics today. CytoCom is a plug-in for Cytoscape to visualise, query and analyse disease-disease networks.

Data set

We collected statistically significant pairwise comorbidity associations reconstructed from over 32 million medical records in the US Medicare claims database recorded in the ICD-9 format (`http://www.icd9data.com`), which are frequently used for epidemiological and demographic studies and collected from Hidalgo et al. (2009). We used MedPAR records from 1990 to 1993, where the dates and reasons for all hospitalisations were reported in ICD-9-CM format and it contains the diagnoses of 13,039,018 elderly patients. In total, the ICD-9 classification consists of 657 different categories at the 3 digit level Hidalgo et al. (2009). Yet, the data set is large enough to predicate race and gender specific comorbidity patterns.

Control Panel

After launching Cytoscape, CytoCom will appear in the control panel and will present itself as shown in Figure 2.

In order to build a disease-disease association network, we need to load the data according to the following steps:

1. Gender selection. Select the "Male", "Female" or "Both" options from the database.
2. Race selection. Select the "White", "Black" or "Both" options from the database.
3. Click the "Load" button, the CytoCom will be connected to the inner database from which the data will be loaded.

After loading the data, we can explore disease-disease association (comorbidity) network by means of the following steps:

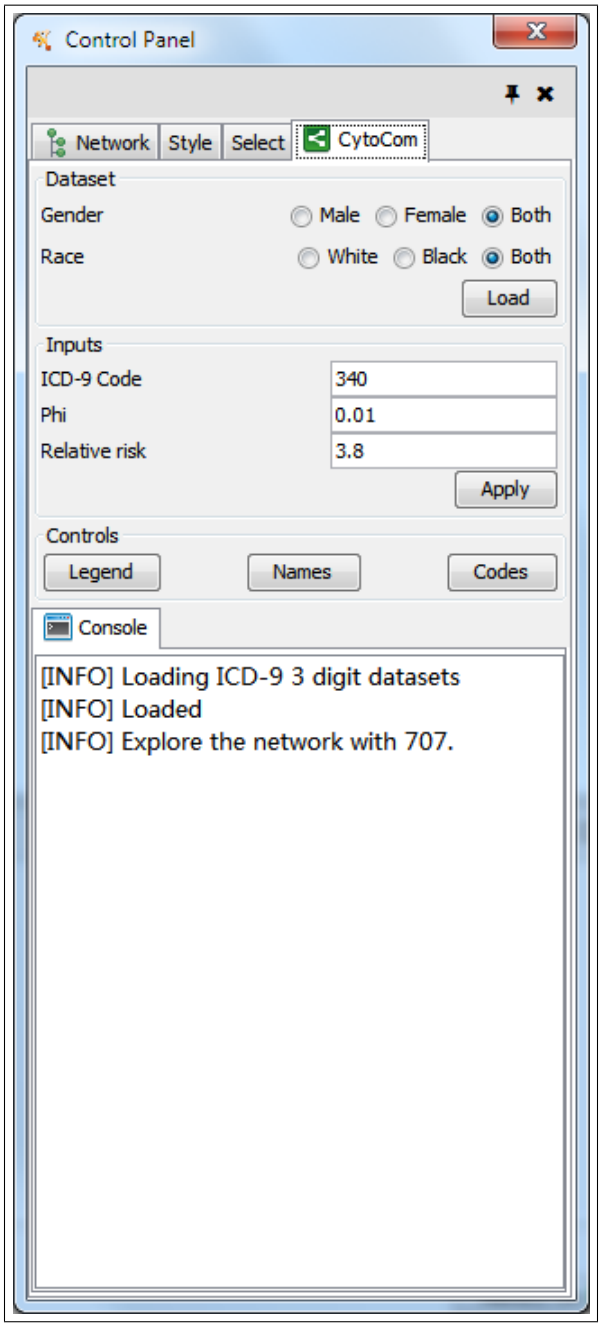


Figure 1. CytoCom control panel

1. Enter the phi values in the text fields. It could be either integer or decimal number. Hence, values greater than input value will be selected to explore the network.

2. Enter the relative risk values in the text fields.It could be either integer or decimal number. Hence, values greater than input value will be selected to explore the network.

3. Click the "Apply" button. The CytoCom will then construct the disease-disease network diagram. The name of the network starts with "Comorbidity network" + "ICD-codes" + "Phi" + "RR". See Figure 2

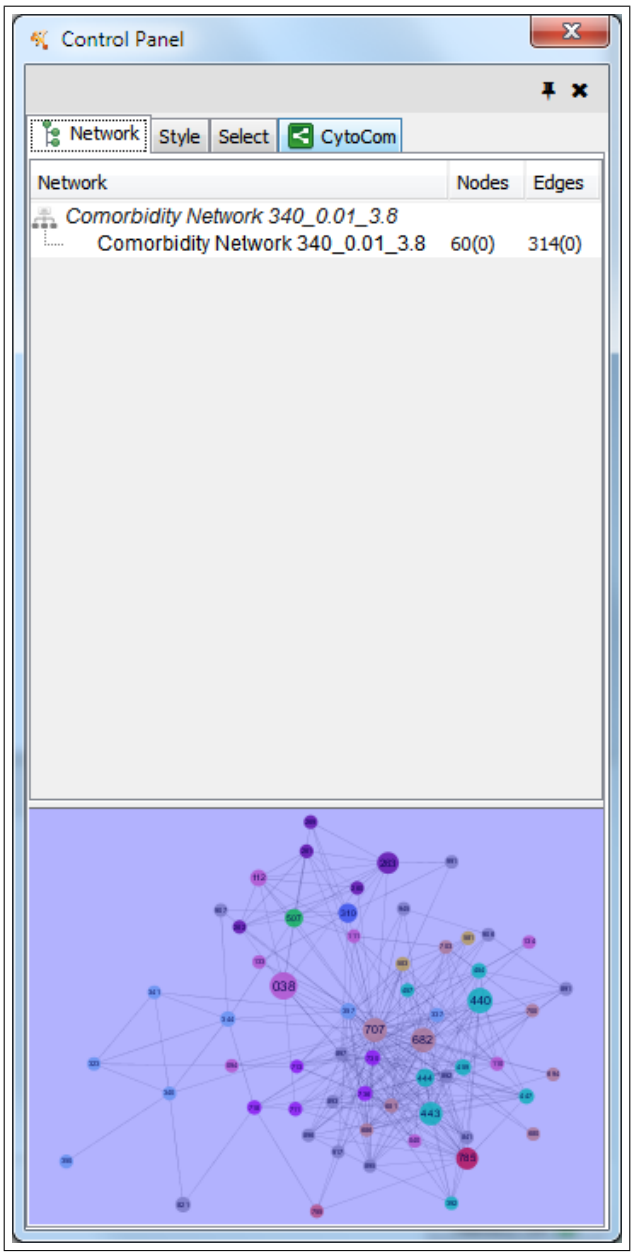


Figure 2. The name of the network is "Comorbidity network 340 0.01 0.6"

In order to explore more diseases that are associated with the disease nodes as seed input in an existing network, users need to search the network by double-clicking the selected nodes within it. A search is then performed based on these nodes using the pre-specified Phi and RR values and the additional comorbid diseases are then added to the existing network. This function works with one or more selected nodes and supports the creation of dynamic network diagrams interactively with CytoCom. Figure 4a shows the network generated with CytoCom using disease 340 and Figure 4b shows the extended disease disease association network starting from the input disease code 350.

Parameter setting

For a pair of diseases i and j, we used two statistical measures to quantify the relationship between two diseases: Relative Risk (RR_{ij}) and ϕ-correlation (ϕ_{ij}), which are calculated based on the Hidalgo et al. (2009). The correlation of $RR_{ij} = 1$ implies no comorbidity, $RR_{ij} > 1$ implies positive comorbidity, and $0 < RR_{ij} < 1$ implies negative comorbidity. Similarly, $\phi_{ij} = 0$ implies no co-morbidity, $0 < \phi_{ij} < 1$ implies positive comorbidity and $-1 < \phi_{ij} < 0$ implies negative comorbidity. The two comorbidity measures are not completely independent of each other, and both measures have their intrinsic biases Hidalgo et al. (2009). They increase with the number of patients affected by both diseases. For example, RR overestimates relationships involving among rare diseases and underestimates the comorbidity between highly prevalent illnesses, whereas ϕ accurately discriminates comorbidities between pairs of diseases of similar prevalence but underestimates the comorbidity between rare and common diseases. Hidalgo et al. (2009) suggested that two diseases are strongly associated if $R_{ij} > 20$ and $\phi_{ij} > 0.06$. Therefore, user may consider these values for relative risk and ϕ to estimate the comorbity among diseases. However, user may observer the disease associations by putting their desired parameters values.

Network construction

The comorbidity disease network is constructed using the two comorbidity measures (relative risk and phi-correlation), and using the patient medical records. All informations that are provided as a data source file is used for the identification and calculation of the comorbidity association between diseases according to the parameter setting by the users. The table panel in Figure 3 provides the representation of the comorbidity output network information. Table contains 8 columns. disease 1 and disease 2 headed columns indicate ICD-9 code of disease 1 and disease 2 respectively. Column name 1

and name 2 represent names of disease 1 and disease 2. Column prevalence 1 and prevalence 2 mean the prevalence of diseases. Both of the prevalence values are the absolute number of affected patients from the specific population. Columns rr and phi are two statistical measures. CytoCom built disease comorbidity network based on these information.

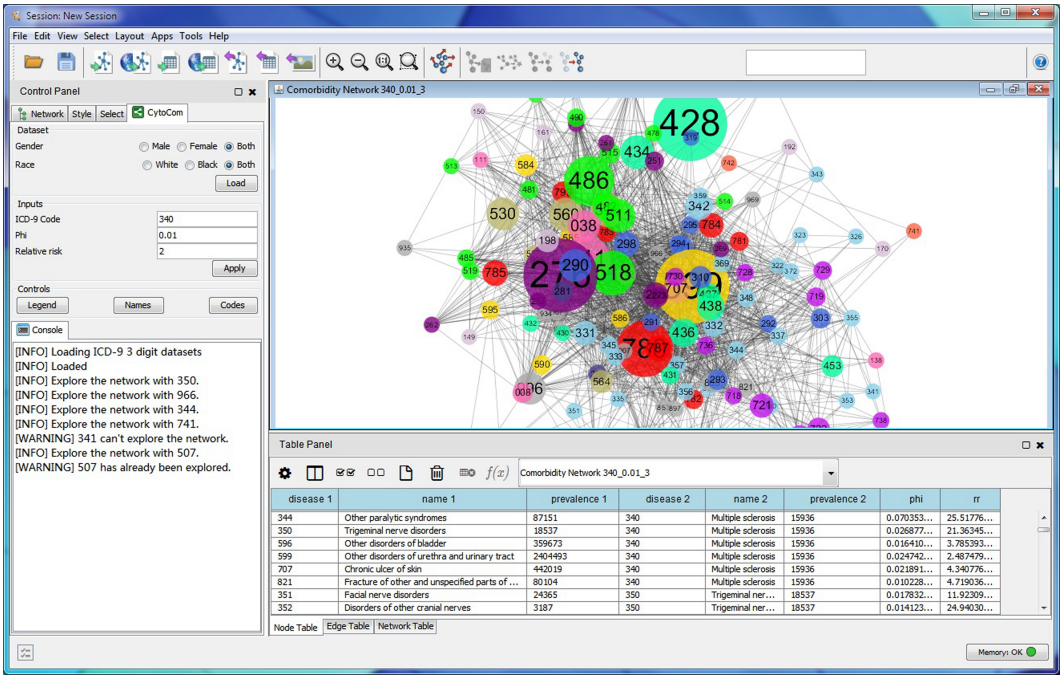


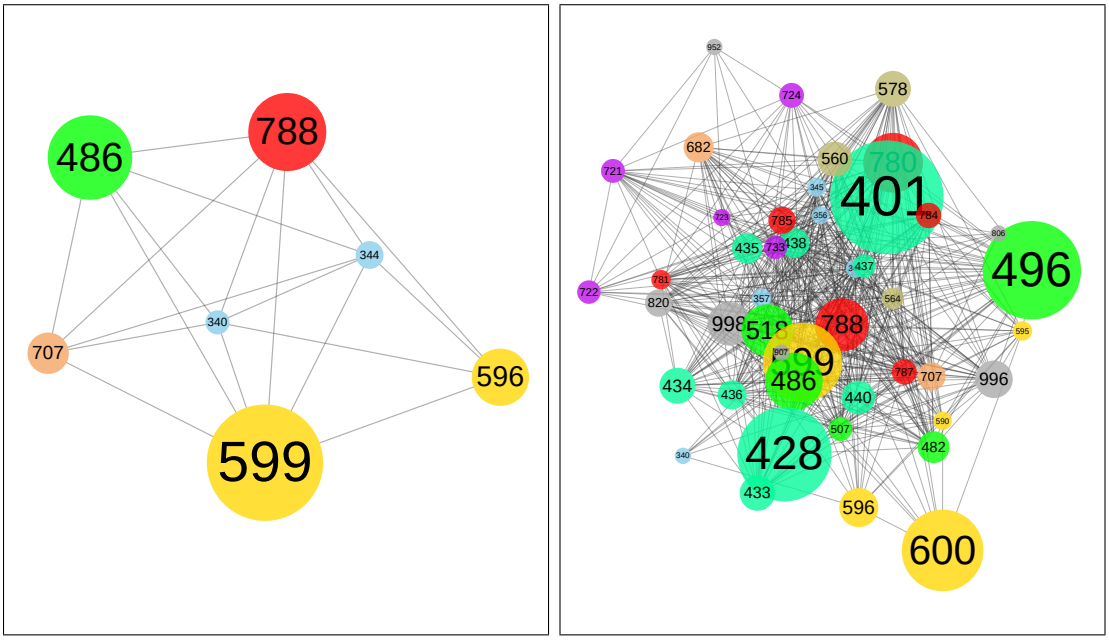
Figure 3. Cytoscape screenshot of CytoCom

Network Visualisation

Each node in the network represents a unique disease. Two diseases are connected if there is an association between them. We colour each node according to the category of diseases based on the first 3 digits of the given ICD9 codes. The node size increases with the increasing the disease prevalence. More common diseases are represented as larger nodes as shown in Figure 4a. In addition, the user can retrieve a display of the legend by clicking the "Show legend" button. It opens the Legend window, which provides colour information for each node, as shown in Figure 5. All nodes can be labelled with either the ICD-9-CM code or the disease name and the "Show names" or "Show codes" tabs can be clicked alternately to switch between labelling the nodes by code or by name.

Customise the network

Cytoscape provides several layout algorithms for organising its network visualisation system. For example, by selecting "Layouts" in the Cytoscape



a b

Figure 4. Panels (a) and (b) show the disease-disease network. Panel (a) shows the network generated with CytoCom using disease 340 and Panel (b) shows the diseases associated with 344 in the existing network that is expended.

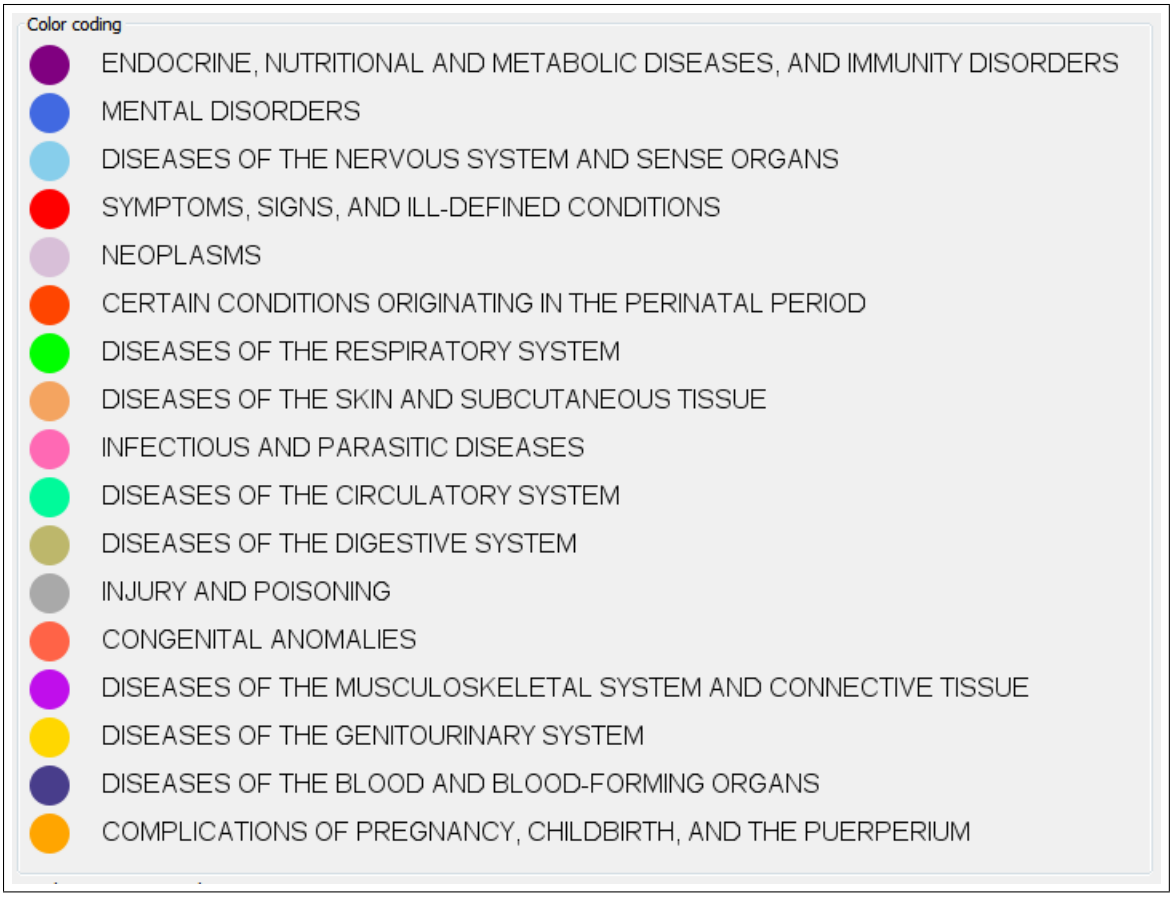


Figure 5. Legend window

menu, it is possible to apply a specific layout algorithm to generate a particular view according to the user's preferred choice. In addition, the user can change the size of the existing network nodes and the width of the existing network edges by using the Style properties of the Cytoscape .

References

Hidalgo, C. A., Blumm, N., Barabási, A.-L., and Christakis, N. A. (2009). A dynamic network approach for the study of human phenotypes. PLoS computational biology, 5(4), e1000353.